

Image to Text and Speech Conversion

Nishu Kumari

Abstract:- This text-to-image convertor aims to check the conversion of data between the various modalities (text, image) because of the evolution of human-machine communication that introduced the utilization of natural communication modalities to humans. Such as gestures, speech, sound, and vision. In fact, one of the main challenges of this "multimodal" learning is the learning of a shared illustration between the distinct modalities and the prediction of the missing knowledge (by retrieval or synthesis) from one conditioned modality to another. Some researchers work on the various varieties of conversions; Text to Speech, Speech to image or Text to image synthesis, and vice-versa however in this paper we tend to can focus on: image to audio image-to-text synthesis.

I. INTRODUCTION

Improve the interface between man and machine in varied Textual info is on the market in several resources like documents, newspapers, faxes, written info, written notes, etc. many folks merely scan the document to store the data within the computers. Once a document is scanned with a scanner, it's kept within the kind of pictures. however these pictures are not editable and it's troublesome to search out what the user requires as they're going to got to undergo the entire image, reading every line and word to work out if it's relevant to their need. pictures additionally take up extra space than word files on the pc. it's essential to be ready to store this information in such the simplest way so it becomes easier to go looking and edit the information. there's a growing demand for applications that will acknowledge characters from scanned documents or captured pictures and build them editable and easily accessible.

Character recognition is one of the foremost interesting areas of pattern recognition and artificial intelligence. Optical Character Recognition extracts the relevant info and mechanically enters it into an electronic information service rather than the traditional way of manually retyping the text. Optical Character Recognition may be a Brobdingnagian field with a variety of various applications like invoice imaging, legal trade, banking, health care trade, etc. OCR is additionally widely used in several alternative fields like Captcha, Institutional repositories and digital libraries, Optical Music Recognition with no human correction or human effort, Automatic variety plate recognition, and Handwritten Recognition. It contributes immensely to the advancement of the associate degree automation process and may applications. Several research works are that specialize in new techniques and strategies that may cut back the processing time whereas providing higher recognition accuracy. currently, it's potential to scan documents as associate degree images and to form editable and searchable for further informatics.

II. OBJECTIVES

The objective of OCR software package aims to organize the text and convert it to editable form. Thus, developing pc algorithms to identify the character within the text's character is OCR. A document is 1st scanned by an associate degree optical scanner, which produces a picture style of it that's not editable. Optical character recognition involves the Translation of this text image into editable character codes like code.

The below diagram shows the process mechanism of OCR system:

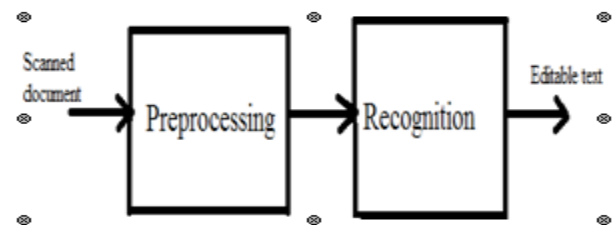


Fig. 1: OCR Engine

III. LITERATURE SURVEY

A. PREVIOUSLY PROPOSED MODEL

In the current world, there's a growing demand for users to convert written documents into electronic documents for maintaining the safety of their information. Hence, the basic Text recognition system was fictional to acknowledge and convert the information available on paper into laptop-processable documents, so the documents will be editable and reusable.

The prevailing system/previous system of Text recognition on a grid infrastructure is simply a text recognition system while not grid practicality. The existing system deals with the same character recognition or character recognition of a single language. The drawback within the early text recognition system is that they solely can acknowledge and convert solely photographs of English or solely of a single language. that the older Text recognition system is Unilingual.

B. PROPOSED MODEL

This planned system is that the Extraction of text from an image recognizer OCR, on a grid infrastructure which will be a personality recognition system that supports the popularity of the characters of multiple languages. This feature is what we have a tendency to tend to call grid infrastructure that eliminates the matter of heterogeneous character recognition and supports multiple functionalities to be performed on the image. throughout this context, Grid infrastructure suggests that the infrastructure supports a cluster of specific sets of languages.

Thus, the Extraction of text from image exploitation of formal logic on a grid infrastructure is multi-lingual.

The advantage of a planned system that overcomes the advantage of the current system is that it supports multiple functionalities sort of a piece of writing and the continuation the text to the character. It to boot adds profit by providing heterogeneous characters recognition. this technique acknowledges the characters supported by the trained info values.

Working Of OCR Model

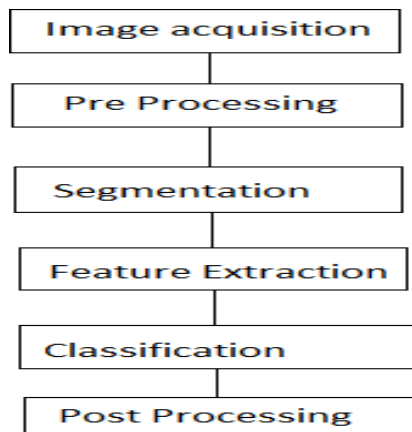


Fig. 2: Working of OCR

A. IMAGE ACQUISITION

The general aim of Image Acquisition is to remodel an associate degree optical image (Real World Data) into an associate degree array of numerical information that may be later manipulated on a pc, before any video or image process will begin a picture should be captured by the camera and reborn into a manageable entity.

B. IMAGE PREPROCESSING

Image preprocessing is that the steps taken to format pictures before they're utilized by model coaching and abstract thought. This includes, however, isn't restricted to, resizing, orienting, and color corrections.

C. SEGMENTATION

Image segmentation involves changing a picture into a set of regions of pixels that area unit delineated by a mask or a labeled image. By dividing a picture into segments, you'll be able to method solely the vital segments of the image rather than process the complete image.

D. FEATURE EXTRACTION

In computer vision and image process, a feature could be a piece of data concerning the content of an image; typically about whether a certain region of the image has certain properties. Features could also be specific structures within the image like points, edges, or objects.

Feature extraction refers to the method of reworking raw data raw information to numerical options which will be processed whereas conserving the data within the original data set. It yields higher results than applying machine learning to the data.

E. CLASSIFICATION

Image classification is that the method of categorizing and labeling teams of pixels or vectors at intervals in a picture supported by specific rules. The categorization law is devised mistreatment one or additional spectral or textural characteristics. Two general ways of classification are 'supervised' and 'unsupervised'.

F. POST-PROCESSING

Post-processing is that the use of any technique or technology to boost the initial image captured by the artist. An old post-processing technique was airbrushing, which was done to get rid of or soften one thing within the original image.

IV. CONCLUSION

Nowadays, applications need several types of pictures as sources of data for elucidation and analysis. once a picture is reworked from one kind to a different one like digitizing, scanning, human activity, storing, etc. degradation happens. Therefore, the output image has got to undertake a method known as image improvement, which contains a group of strategies that request to develop the visual presence of a picture. Image improvement is enlightening the interpretability or awareness of data in pictures for human listeners and providing higher input for different automatic image process systems. OCR image process may be a powerful tool for kind preparation of knowledgeable information edge and therefore the combination of inaccurate data from totally different sources. The supposed tesseract rules square measure a horny result to enhance the standard of edges the maximum amount as potential.

ACKNOWLEDGMENT

The authors wish to thank Er. Amandeep Kaur and Er. Mariam Khan for their guidance.

REFERENCES

- [1.] L. Neumann and J. Matas. A method for text localization and recognition in real-world images. In ACCV, pages 770–783, 2010.
- [2.] Reza Sarshogh and Keegan Hines, "Computer Vision Methods for Extracting Text from Images", Capital One Tech.
- [3.] T. Som, Sumit Saha, "Handwritten Character Recognition Using Fuzzy Membership Function", International Journal of Emerging Technologies in Sciences and Engineering, Volume 5, December 2011. L. A. Zadeh. Fuzzy sets, Information Control 8 (1965) 338–353.
- [4.] Gur, Eran, and Zeev Zelavsky, "Retrieval of Rashi Semi-Cursive Handwriting via Fuzzy Logic," IEEE International Conference on Frontiers in Handwriting Recognition (ICFHR), 2012
- [5.] Thomas Natsvhlager, "Optical Character Recognition", A Tutorial for the Course Computational Intelligence.
- [6.] Andrei Polzounov, Artsiom Ablavatski, Sergio Escalera, Shijian Lu, Jianfei Cai "Wordfence: Text

- Detection In Natural Images With Border Awareness”
- [7.] D. Trier ,A.K.Jain ,TText , Feature Extraction Method for Charac- ter Recognition-A Survey” , Pattern Recognition
- [8.] https://en.wikipedia.org/wiki/Optical_character_recognition
- [9.] Image to Text Converter
Website <https://www.prepostseo.com/image-to-text>